2001

Cloning Of HsdS Gene and Characterization of the Protein From Haemophilus Influenzae Rd

Srividya Chalapatti
Seton Hall University

CLONING OF *hsd*S GENE AND CHARACTERIZATION OF THE PROTEIN FROM *HAEMOPHILUS INFLUENZAE* Rd

Srividya Chalapatti

Approved By

Maria P. MacWilliams

DR. MARIA P. MACWILLIAMS
MENTOR

Marian Glenn

DR. MARIAN GLENN
COMMITTEE MEMBER

DR. SULIE CHANG
COMMITTEE MEMBER

DR. SULIE CHANG
CHAIRPERSON, BIOLOGY DEPARTMENT

Acknowledgements

I would like to thank Dr. Maria P. Macwilliams, my mentor, for help and advising in my research. I would also like to thank Dr. Mark Chiu, for his help and expertise with Protein isolation and SDS-PAGE gels.

Finally I would like to acknowledge my husband, Murali Talanayar for all the support and help in scanning pictures and my parents, C.S. Jayaramasundaram and C.S. Bharathi, for their wishes and blessings in completing my thesis.

TABLE OF CONTENTS

TITLE	1
APPROVAL PAGE	2
ACKNOWLEDGEMENTS	3
TABLES	5
FIGURES	6
ABSTRACT	7
INTRODUCTION	8-14
MATERIALS AND METHODS	15-28
RESULTS	29-46
DISCUSSION	47-50
LITERATURE CITED	51-52

TABLES

	TITLE	PAGE
I	Examples of DNA sequences recognized by the type I enzymes	5
II	Putative RM systems in sequenced genomes	8
III	Bacterial Strains	11
IV	Plasmids	11
V	Plasmids constructed in the present work	11
VI	EOP values of restriction assays using LMG194 cells containing pSRI4 and / or pHJ8	30
VII	Restriction assay EOP values using LMG194 cells containing pSRI5 and / or pMG3	31

FIGURES

	TITLE	PAGE
1	The type I restriction endonuclease subunits	5
2	Model of DNA restriction by type I RM enzymes	6
3	PCR primers used to amplify the 5'end of HI0216 gene	12
4A	Cloning scheme-pSRI1	14
4B	Subcloning- pSRI4	15
4C	Cloning of pSRI5	17
5	Agarose gel electrophoresis of PCR product	25
6	pSRI1 isolates digested with *Ssp*I to confirm PCR insert	25
7	*Eco*RI digestion of pSRI4 isolates to confirm orientation	26
7A	pSRI4 map showing the *hsd*S gene insert from H.influenzae	27
8	Plasmid map of pSRI5 showing *hsd*S and *hsd*M inserts	28
9	Confirmation of pSRI5 inserts by digestion with *Bsp*DI/*Cla*I	28
10	Protocol for *in vivo* restriction assay	29
11	Protein expression induction at various arabinose concentrations	32
12	Analysis of HsdS expression at different time points in induced and uninduced cultures of LMG194/pSRI5	33
13	Optimization of protein expression using same cell mass of LMG194/pSRI5	34
14	SDS PAGE analysis of DEAE column fractions from peak of protein elution	35
15	SDS PAGE analysis of Heparin agarose column fractions from peak of protein elution	36
16	Heparin agarose protein fractions 14-29 on 12.5% SDS PAGE	37
17	Confirmation of *hsd*M by western blotting technique using ECL detection kit	39
18	*In vitro* DNA methylation assay	41
19	Amino acid sequence comparison between HI0216 and *Eco*DXXI *hsd*S gene	43

Abstract

Restriction-Modification (R-M) systems serve as a defense against foreign DNA and viruses. The two main activities of all R-M systems are methylation of host DNA at a specifically recognized DNA sequence and double strand cleavage of unmodified DNA. The diverse classes of R-M systems are based on enzyme subunit composition, reaction cofactors and mechanism of DNA cleavage. In type I R-M systems both the DNA methylation and restriction reactions are catalyzed by a complex of three subunits: *hsd*S, *hsd*M and *hsd*R. *hsd*S is the DNA binding specificity subunit, *hsd*M catalyzes methylation and *hsd*R is responsible for DNA restriction. A further subdivision of type I systems into families groups closely related enzymes whose subunits are interchangeable.

Genomic sequencing has revealed the presence of putative type I genes in diverse species, both bacterial and archeal. To characterize one alleged *hsdS* gene from *Haemophilus influenzae Rd*, the HI0216 ORF was cloned behind the arabinose-inducible promoter, P_{BAD}. The resulting plasmid, pSRI4, was transformed into *E. coli* cells along with a compatible plasmid which expresses type IC *hsd*M and *hsd*R genes constitutively. The cells produced a functional restriction endonuclease as measured by *in vivo* restriction of bacteriophage lambda. Restriction required an arabinose preincubation which demonstrated that the HI0216 gene product was required for activity. Thus HI0216 interacts with the enteric-derived *hsd*M and *hsd*R subunits to assemble a functional enzyme. These results demonstrate HI0216 *hsd*S gene to be a member of the type IC family.

Introduction

The phenomenon of restriction and modification (R-M) of foreign DNA was first observed in studies on bacteriophages. Bacteria in their natural environment are faced with predation by both macro and microorganisms. Some of the most important predators are the bacteriophages. The host bacteria have evolved a number of means to prevent destruction by the phages (Snyder, 1995) (Klaenhammer, 1987; Klaenhammer, 1989). One effective means of defense against viral attackers is to prevent any productive contact between the phage and bacteria. This is done by secretion of barrier (ie. capsule layer) that prevents the approach of the phage or by mutation of genes encoding phage receptors present in cell membrane. Another stage at which bacteria can fend off viral attackers is the injection of the viral DNA into the host cell. R-M systems function at this point to inhibit viral multiplication. The classical mechanism of restriction and modification of bacteriophages consists of endonucleolytic cleavage or restriction of phage DNA when it is not specifically methylated or modified at certain sequences (Bickle and Krüger, 1993); (Bickle, 1987); (Boyer, 1971).

R-M systems are found exclusively in unicellular organisms, mainly bacteria, and in some viruses. Their main purpose is the protection of a cell from invading foreign DNA, although it is hypothesized that they play a role in acquisition of new DNA sequences (Arber, 1990; Arber, 1994). A prerequisite to both the functions is the ability to distinguish between foreign DNA and the cell's own genome. R-M systems consist of two opposing DNA-sequence specific enzyme activities: a restriction endonuclease that cleaves the DNA unless the recognition sequence is methylated and protected by the corresponding

modification methyltransferase. Invading foreign DNA lacking the modification will be cleaved and subsequently degraded. R-M systems are classified into three groups: types I, II and III based on the complexity of enzyme structure and cofactor requirements.

Type II systems are the best characterized of the R-M systems. They have separate restriction and modification enzymes that act independently of each other and have simple cofactor requirements. Restriction depends on Mg^{2+} while modification depends on S-Adenosyl-L-Methionine (Adomet). The type II restriction enzymes cleave DNA at fixed positions within or at a defined distance from their recognition sequences. The exquisite cleavage specificity and simple cofactor requirements of the type II endonucleases have made them essential tools in molecular biology.

The type III class contains the fewest members. The restriction enzyme contains two subunits: Mod and Res. Mod can function alone as a DNA methyltransferase and requires Adomet as the methyl group donor. The Res subunit has no enzymatic activity when not complexed with Mod. Restriction requires Mg^{2+} and ATP, which is hydrolyzed at a low rate, the cleavage reaction is stimulated by Adomet.

Our research focuses on the type I R-M systems which are the most complex of the three classes. The type I enzymes can function as restriction endonucleases, ATPases or DNA methyltransferases. The restriction endonuclease cleaves unmodified DNA in presence of Mg^{2+}, ATP and Adomet. Only the methyl donor is required in the modification reaction. (Bickle, 1993). They recognize asymmetric sequences consisting of two half sites, one of 3 or 4 basepairs and other of 4 or 5 basepairs which are separated by a nonspecific spacer of 6-8 basepairs. (Table I).

Table I: Examples of DNA sequences recognized by the Type I enzymes
(Adapted from Bickle, 1987)

Enzyme	Sequence Recognized
EcoB	TGANNNNNNNNTGCT ACTNNNNNNNNACGA
EcoK	AACNNNNNNGTGC TTGNNNNNNCACG
EcoDXXI	TCANNNNNNNNATTC AGTNNNNNNNNTAAG
EcoR124	GAANNNNNNNRTCG CTTNNNNNNNYAGC

N= Any nucleotide; R= either Purine, Y= either Pyrimidine. The adenine residues methylated in modification reaction are marked with asterisks.

These enzymes methylate adenine residues, one in each component of the target sequence but on opposite strands.A type I system is specified by three contiguous genes organized into two transcriptional units. The three genes *hsd*S, *hsd*R and *hsd*M code for the three subunits of the multifunctional restriction endonuclease HsdS, HsdR and HsdM (Figure1). (Bickle, 1993)

Figure 1. The Type I Restriction endonuclease subunits
The type I restriction endonuclease is composed of three nonidentical subunits: HsdR, HsdM, HsdS. The tetraaminoacid repeat $(TEAL)_3$ present in type IC enzymes is indicated.

The modification methyltransferase contains only the *hsd*S and *hsd*M gene products while the restriction endonuclease contains HsdR in addition to HsdS and HsdM. The HsdR subunit is required for restriction, the HsdM subunit is necessary for modification and the HsdS subunit is responsible for specific DNA recognition in both modification and restriction reactions. The restriction endonuclease can catalyze both the DNA restriction and methylation reactions.

When the recognition site is methylated in both strands the DNA containing the site is not a substrate for the R-M complex. A hemimethylated recognition site will be methylated on the unmethylated adenine. When both adenine residues are unmethylated the DNA is cut. Restriction of linear DNA requires two unmethylated sites (Figure 2). Cleavage is nonspecific. Binding of the enzyme to the unmodified recognition site activates enzyme for restriction. ATP hydrolysis fuels translocation of DNA past the enzyme. DNA cleavage is elicited by collision between two translocating enzymes.

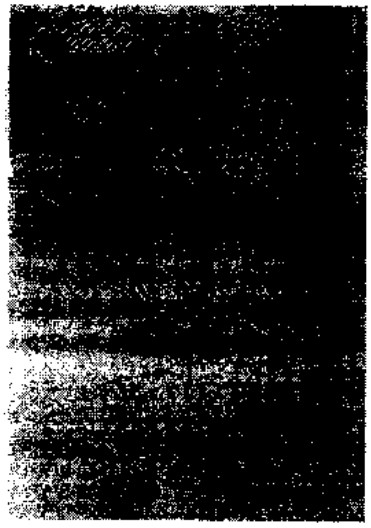


Figure 2. Model of DNA restriction by type I R-M enzymes.
Two recognition sequences (arrowheads) are occupied by one endonuclease molecule each (hatched ovals). The unmethylated state of the recognition site stimulates DNA translocation passed the endonuclease which remains bound to the recognition site. When two oppositely translocating endonucleases meet, a conformational change occurs (grey ovals) activating them for restriction (black arrow). After cleavage the products become methylated (Figure from Dreier *et al,* 1996).

Type I systems are subdivided into families, A, B, C and D based on sequence similarity, antigenic cross-reactivity and genetic complementation between the subunit genes. Some members of type IC family are plasmid coded. The best known IA, IB and ID systems occupy the same chromosomal location in different strains of *E. coli* and thus can be considered allelic. Within each family the *hsd*R and *hsd*M genes have high levels of sequence similarity and are exchangeable. The HsdS subunits of a family share conserved aminoacid regions that are thought to be responsible for protein-protein interactions with the HsdM and HsdR subunits. The two extensive variable regions present in HsdS subunits are proposed to be sequence-specific DNA recognition domains.

Type IC HsdS subunit structure is ideal for generation of new DNA recognition specificities: the two variable DNA binding domains are flanked by the regions conserved amongst all family members. Hybrid systems with new binding specificities can be generated by recombination between two IC enteric systems within the DNA sequences coding for the central, conserved region. In addition, a repeated DNA sequence encodes a tetra aminoacid repeat in the central conserved region. The number of repeats governs the length of the nonspecific spacer in recognition sequence: three repeats results in a 6 bp spacer and four repeats a 7 bp spacer.

From a molecular point, the cloning and characterization of the type I systems is interesting because of the highly variable DNA binding domains each of which recognizes a unique D NA sequence. The great diversity of DNA binding domains may lead to a better understanding of how these domains recognize DNA. The facility with which these domains can be swapped also makes these systems fascinating models for gene evolution.

Type I enzymes were originally found in members of the Enterobacteriaceae such as *E. coli* and *S. typhimurium*. However type I systems are not confined to enterics alone but are present in many diverse bacteria. The widespread distribution of these enzymes has been revealed through the recent genome sequencing projects (Table II).

Organism	Type I	Type II	Type III
Escherichia coli K12	1	-	-
Haemophilus influenzae Rd	2	3	1
Mycoplasma genitalium	-		-
Mycoplasma pneumoniae	1		-
Helicobacter pylori	3	4	4
Methanococcus jannaschii	3	2	
Methanobacterium thermoautotrophicum	1	-	-
Archaeolobus fulgidus	1	-	1
Borrelia bugdorferi	-	-	-

Table II. Putative R-M systems in sequenced genomes, (partial list. www.tigr.org)

The *H. influenzae* Rd genome sequence (Genome Sequence Database Accession Number L42023) represents the first complete genome sequence from a free-living organism. Determination of entire genome sequence is the prerequisite to understanding of the complete biology of the organism. *H. influenzae* is a small, nonmotile, gram-negative bacterium whose only natural host is human. They are commensal residents of upper respiratory mucosa of children and adults and cause respiratory tract infections, mostly in children. This particular strain of *H.influenzae* has great historic significance to the field of molecular biology because it served as the source of the first type II restriction endonuclease, *Hin*DII and *Hin*dDIII.

Type I RM systems are uniquely suited for the diversification of sequence specificity. In part this is because a single subunit is responsible for the sequence specificity of both enzyme activities. In addition, some, perhaps all type I RM systems have coevolved with their hosts to ensure that neither the acquisition of such a system nor a change in specificity leads to endonucleolytic cleavage of the host chromosome. The scope of diversification is enhanced by presence of two different TRD's within one specificity subunit.

RM systems therefore influence the flux of genetic material between bacterial populations, enhancing the opportunity for the acquisition of advantageous coding sequences in the absence of deleterious ones. The organization of type I enzymes, the bipartite asymmetrical nature of target sequence, and host-mediated alleviation of restriction emphasize the importance of diversity.

We are interested in the diversity of the type IC enzymes. My research project compared a putative type I *hsd*S gene to its Gram negative enteric system counterparts. One putative *hsd*S gene from *H. influenzae*, the HI0216 ORF, was cloned and characterized. The resulting construct was analyzed for the production of a functional HsdS subunit that complemented mutants of well-characterized enteric systems.

Materials and Methods

Bacterial strains, bacteriophages and plasmids

The bacterial strains used are listed in Table III. Standard bacterial culture methods used were as described by Sambrook *et al.* (1989). $\lambda_{virulent}$ high titer lysates were prepared by the plate lysate method described by Silhavy, Berman and Enquist(1984). Previously constructed plasmids used in this study are listed in Table IV while those constructed during the present project are listed in Table V. Standard bacterial media (top agar, SM buffer, Luria-Bertani (LB) broth and agar plates) were prepared as described in Sambrook *et al* (1989). When necessary antibiotics were added to LB broth and plates at the following concentrations: Ampicillin 200μg/ml, Chloramphenicol 30μg/ml.

Molecular biology reagents and procedures: Restriction enzymes, T4 DNA polymerase, Vent™ thermostable polymerase, Calf intestinal phosphatase (CIP) were from New England Biolabs. T4 DNA ligase was obtained from Life Technologies. All enzymes were used according to manufacturer's instructions. Geneclean kit for purification of DNA fragments from agarose gels and buffer exchange steps according to manufacturer's instructions (Bio101). Deoxynucleotides from New England Biolabs and Amersham Pharmacia Biotech (APB). The DNA molecular weight markers used were from Life Technologies, Promega Corporation. Protein molecular weight markers were from Sigma or APB. $CaCl_2$ competent cells were prepared and DNA transformations were carried out using a variation of the methods in Sambrook *et al* , 1989. In some cases, DNA was introduced into *E. coli* cells by electroporation using the Biorad Gene Pulser apparatus according to the manufacturer's instructions. DNA agarose gel electrophoresis and DNA Polyacrylamide gel electrophoresis from Sambrook *et al* (1989).

Table III. Bacterial strains

Name	Genotype	Reference
DH5α (R^-M^+)	F^- *end*AI *hsd*R17($r_k^-m_k^+$) *sup*E44 *thi*-1 *rec*A1 *gyr*A (Nal^r) *rel*A1Δ(*lac*ZYA-*arg*F)U169 *deo*R(ϕ80d*lac*Δ(*lac*Z)M15)	Woodcock *et al.* (1989)
LMG194 (R^+M^+)	F^-Δ*lac*X74 *gal*E *thi rps*L Δ*pho*A (*pvu*II) Δ*ara* 714 *leu*::Tn10	Guzman *et al.* (1995)

Table IV. Plasmids

Plasmid	Characteristics	Reference
pUC18:GHIJT88	HI0216 ORF; Ap^R	www.tigr.org
pBAD22	P_{BAD} multiple cloning site; AraC; Ap^R	Guzman et al (1995)
pHJ8	*Eco*DXXI *hsd*M+*Eco*R124 *hsd*R; Cm^R	MacWilliams & Bickle (1996)
pMG3	*Eco*R124I *hsd*R; Cm^R	Gubler & Bickle (1991)
pMMW56	*Eco*DXXI *hsd*M; Ap^R	MacWilliams & Bickle (1996)

Table V. Plasmids constructed in the present work

Plasmid	Characteristics Plasmid size; insert; Antibiotic resistance	Vector; cloning site
pSRI1	1155 bp *hsd*S ; Ap^R	pBAD22, *Nco*I-*Xba*I;
pSRI4	5771 bp ;full length HI0216 *hsd*S Ap^R,	pSRI 1, *Xba*I
pSRI5	10.2 kb ; *Eco*DXXI *hsd*M:HI0216*hsd*S; Ap^R	pSRI4, *Bam*HI/*Xho*I; *Hin*DIII/*Sal*I

Plasmid Construction

PCR amplification and cloning of the 5'end of HI0216

The putative hsdS gene, HI0216, was cloned into the arabinose regulated expression vector, pBAD22 in two steps (figure 4A). The 5' end of the HI0216 gene was PCR amplified (see below) inorder to introduce an *Nco*I restriction site which would overlap the ATG start codon of HI0216. The PCR primers used are listed in figure 3. pUC18:GHIJT88 (Table IV) was used as the template.

5'primer: GCTGAC**CCATGG**AAAACAACCGCACT

3'primer: CGTACCTGTCGAC**TCTAGA**CTAGCAGAAC

Figure 3. PCR primers used to amplify the 5'end of HI0216 gene.
The 5'end of the gene was amplified in order to introduce restriction sites for cloning. The new *Nco*I and *Xba*I sites, unique in the PCR product are shown in bold face type.

For the PCR reaction, 0.5 μmoles of each primer were mixed with 5 ng/μl template DNA in 1x New England Biolabs Thermopolymerase buffer (10 mM KCl, 20 mM Tris-HCl pH 8.8 at 25°C, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Trition-X-100) containing 200 μM each dNTP. 0.5 μl of Vent® polymerase were added and the reaction mixture was overlaid with mineral oil. The amplification reactions were run in a Hybrid Omni Gene Temperature Cycler programmed as follows:

Stage I: Step 1: Denaturation- 92°C, 3 minutes.

Stage II: Cycles:5, Step 1: Denaturation- 92°C, 1 minute. Step 2: Annealing- 40°C, 30 seconds. Step 3: Synthesis- 72°C, 30 seconds

Stage III: Cycles: 20, Step 1: Denaturation- 92°C, 1 minute. Step 2: Annealing- 40°C, 30 seconds. Step 3: Synthesis- 72°C, 30 seconds

The PCR product as well as the pBAD22 vector was digested to completion with *Nco*I and *Sma*I. Both digested DNAs were Gene Cleaned to concentrate the DNAs and the DNAs were resuspended in TE (10mM Tris.HCl, pH 8.0, 1mM EDTA). The digested PCR product was run on a 1.2% agarose gel in 1x TAE and 0.1-0.5μg/ml ethidium bromide. The appropriate sized DNA 230 bp fragment was excised and the DNA was purified using Geneclean (figure 5) The purified PCR product and digested pBAD22 were mixed in varying insert to vector ratios (2:1; 5:1) and ligated overnight at 15°C. Transformation of *E. coli* DH5α, plating on selective medium and screening of miniprep DNA resulting in isolates containing the desired plasmid, pSRI1 (Figure 6).

Subcloning rest of *hsd*S gene

To reconstruct the HI0216 gene, the 3' end of the *hsdS* gene 1 Kb was subcloned into pSRI1 (figure 4B). pUC18: GHIJT88 was digested with *Xba*I, which cuts at the *Xba*I site present in HI0216 and at a second site within the multiple cloning site at the end of the *H. influenzae* insert. This 1Kb *Xba*I fragment was gel purified using the Geneclean kit. pSRI1 was digested with *Xba*I and treated with calf intestinal phosphatase (New England Biolabs) to prevent vector self ligation. The DNAs were mixed and ligated. The resulting plasmids contained the new insert in either of two orientations. Competent DH5α cells were transformed with the ligation mixture. Miniprep DNAs from the resulting Ap^R clones were screened for the correct orientation by digestion with *Eco*RI (figure 7). The plasmid isolates with 0.6 Kb and 5.2Kb *Eco*RI fragments contained the correct orientation plasmid, which reconstructed HI0216. This construct was designated as pSRI4 (figure 7A).

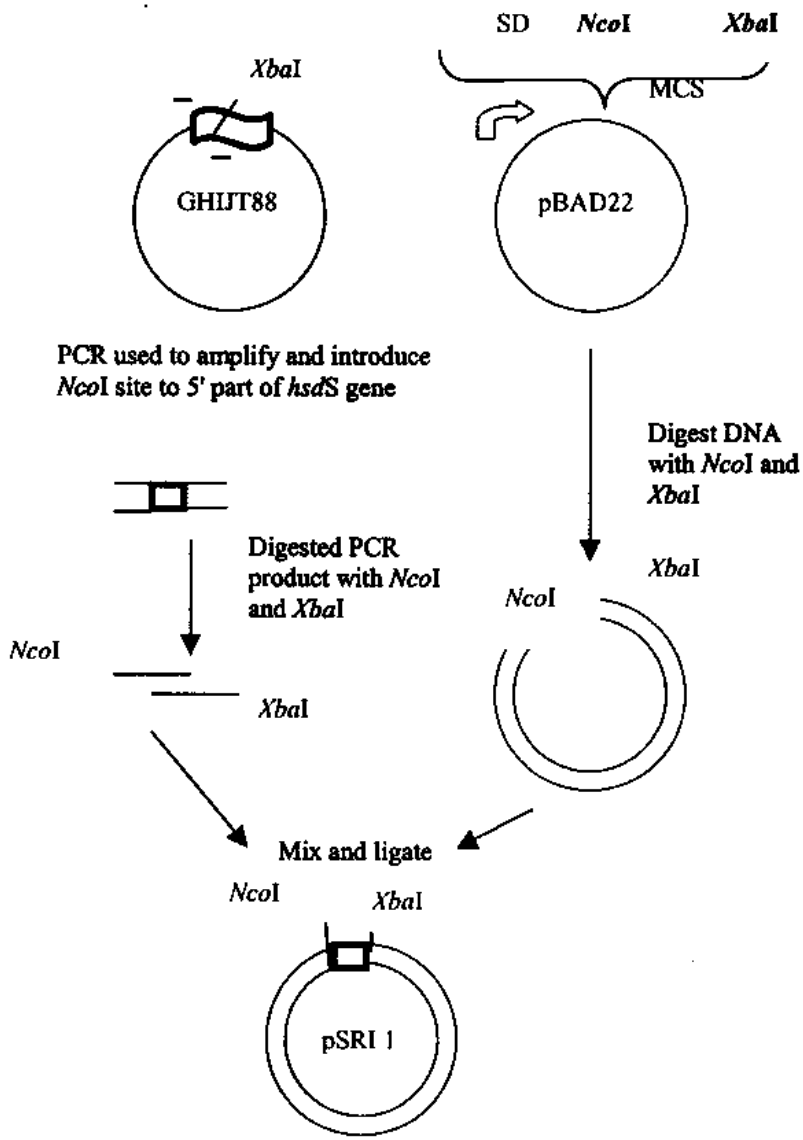


Figure 4A. Cloning scheme to produce pSRI1. Cloning of 5' end of *hsd*S gene from *H. influenzae* into arabinose expression vector pBAD22 resulting in pSRI1 plasmid.

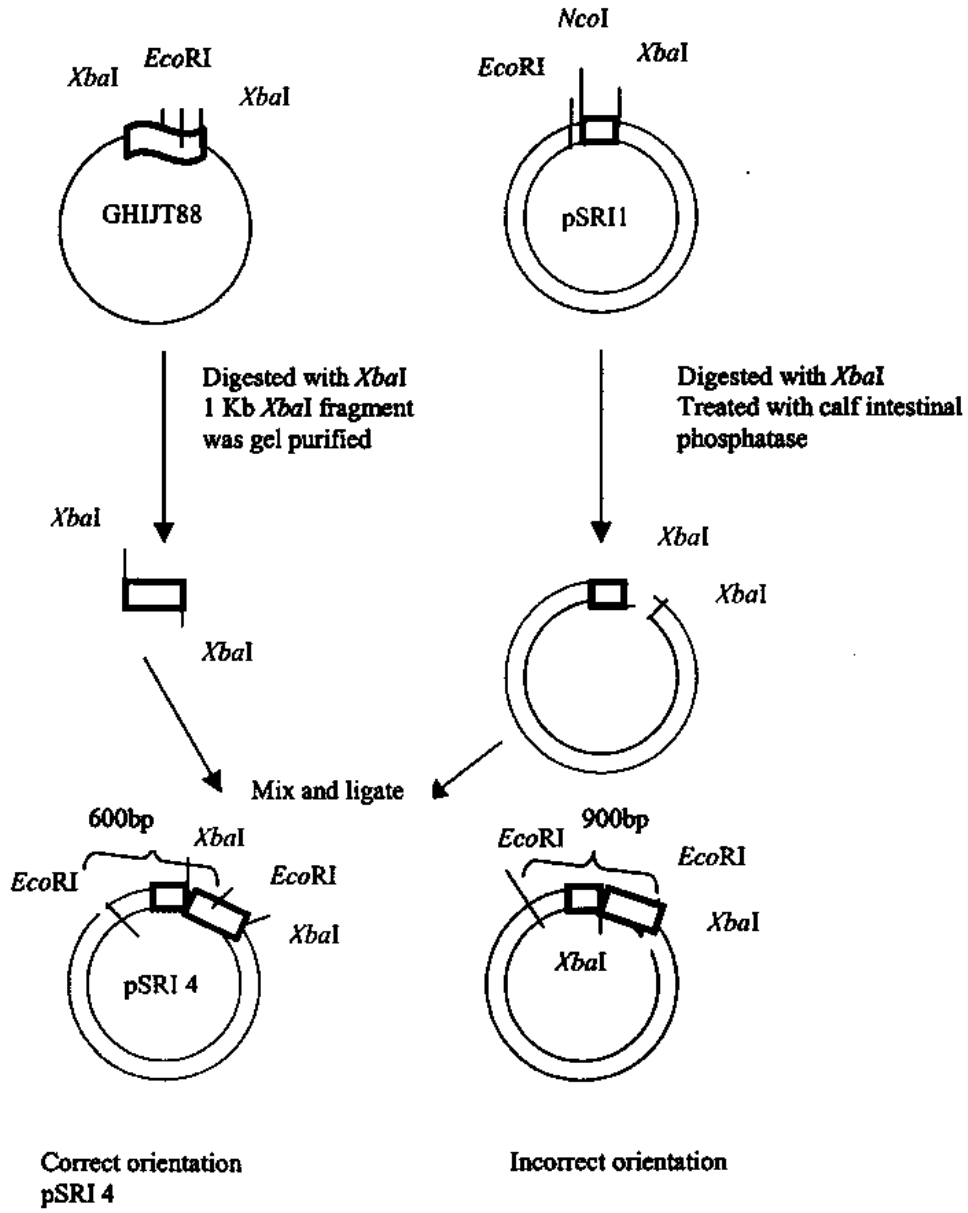


Figure 4B. Subcloning of pSRI4. Subcloning of rest of *hsd*S gene into pSRI1 to construct pSRI4

Cloning of *hsd*M from *Eco*DXXI and *hsd*S from pSRI4 to produce pSRI5

pMMW56 was digested with *Bam*HI, treated with T4 polymerase to fill in the 5' overhangs, and then digested with *Xho*I (figure 4C). This produced two fragments, approximately 4.5 Kb and 3 Kb in length. The digested DNA was run on a 0.8% agarose gel. The 4.5 Kb *hsd*M containing fragment was excised from the gel and further purified using Geneclean. pSRI4 was cut with *Hind*III, treated with T4 polymerase to produce blunt ends and digested with *Sal*I. The digested pSRI4 preparation was then genecleaned . The two DNAs were mixed and ligated to produce pSRI5, which contains both *hsd*S and *hsd*M and is approximately 9.5 Kb (figure 8). To check for the presence of *hsd*M insert, prospective positive DNA clones were digested with *Bsp*DI (isoschizomer of *Cla*I, and run on 1% agarose gel to identify the characteristic DNA fragments of 6.1 Kb and 3.5 Kb (figure 9).

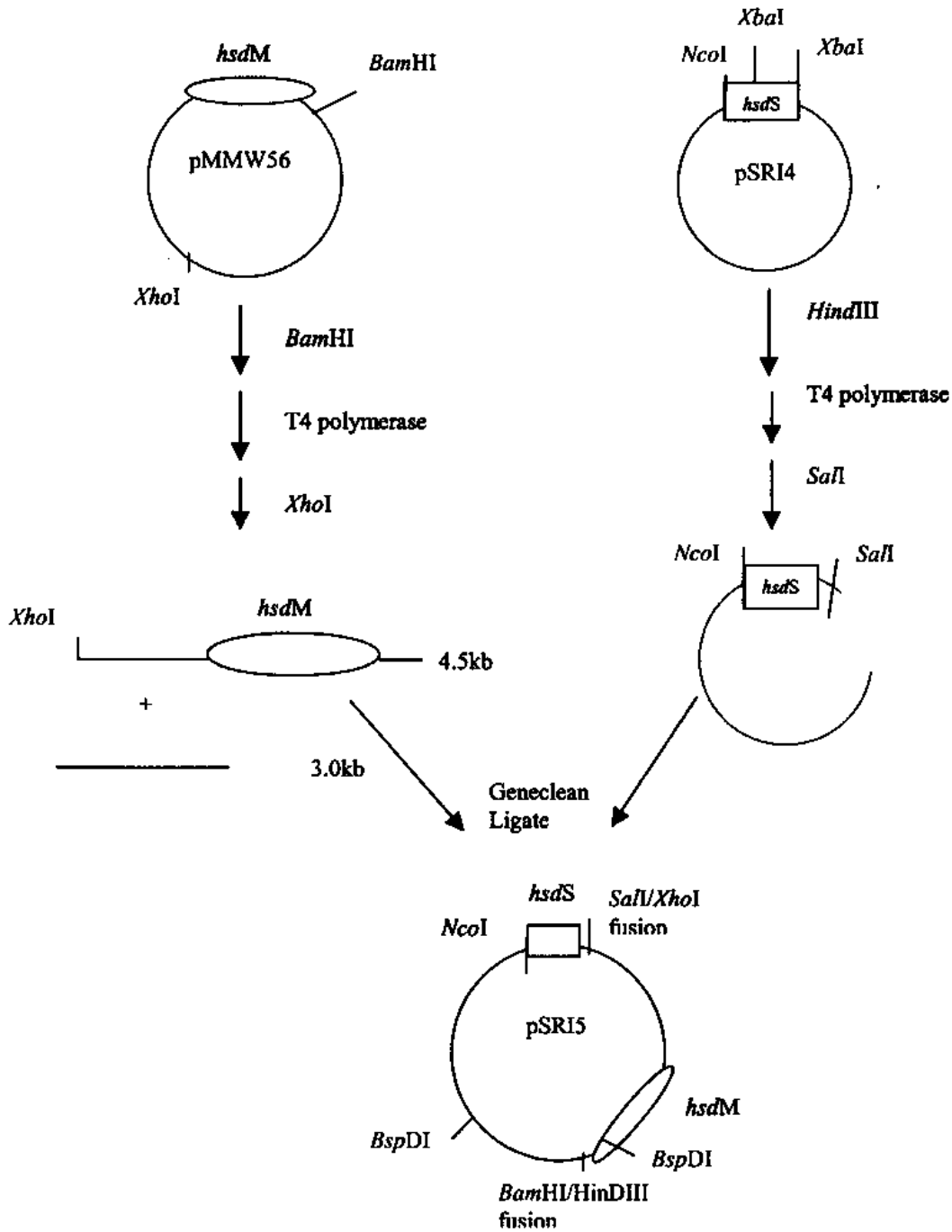


Figure 4C. Cloning of pSRI5. Cloning of *hsdM* from pMMW56 into pSRI4 to construct pSRI5, an expression vector for HI0216 methyltransferase.

Transformation of *E.coli* strain LMG194 with plasmid DNA

In order to test whether or not the HI0216 gene product functioned as the specificity subunit for a type I enzyme, we constructed an *E.coli* strain which contained pSRI4 and a second plasmid, pHJ8 (constitutively expresses the *hsd*M and *hsd*R genes from the enteric Type IC systems). LMG194, a bacterial strain designed for arabinose expression vector use, was transformed with the two plasmids and the resulting Cm^R, Ap^R clones were checked for the presence of the plasmids. To confirm the presence of the inserts miniprep DNA was prepared and digested with *Hind* III which linearized both plasmids. Agarose gel analysis of the digests revealed DNA fragments of the correct size.

***In vivo* restriction assay**[3]

Overnight cultures of LMG194 containing pSRI4, pHJ8 or both plasmids were used. Appropriate antibiotics (Ampicillin 100 μg/ml and Chloramphenicol 30 μg/ml) were added to the medium to select for the plasmids. Serial dilutions of an unmodified phage lambda high titer lysate were prepared. The overnight cultures were diluted 1:50 in LB + selective antibiotics and the subcultures were incubated at 37 °C with agitation. When the cultures reached an OD_{600} of 1.0, 0.2% Arabinose was added to induce *hsd*S expression. After 1 hour incubation, 0.4 ml induced cells was mixed with 0.1ml of a phage dilution and the mixture was incubated for 20 minutes at room temperature. 3 ml top agar was then added to the infected culture; the mixture was vortexed and poured on to LB-Difco plates. After the agar solidified, the plates were incubated at 37°C overnight. The efficiency of plating (EOP), a measure of restriction activity, was calculated by the ratio of the number of plaques/ml of restricting strain by the number of plaques/ml of nonrestricting strains (figure 10).

A similar *in vivo* restriction assay was used to demonstrate a functional restriction activity when the methyltransferase producing plasmid, pSRI5 was coexpressed with the *hsd*R coding plasmid, pMG3 (table VII).

Optimization of protein expression

pSRI5 DNA was transformed into LMG194 cells and the resulting transformants were used in protein expression optimization studies. The two conditions varied were arabinose concentration and induction time length. The parameters utilized were those of the Invitrogen protocol (Invitrogen, 1998). An overnight culture of LMG194/pSRI5 was subcultured 1:100 in 10 ml LB + Ampicillin 100 μg/ml. Five subcultures were prepared and incubated with shaking until the cultures reached mid log phase ($OD_{600} \sim 0.5$). At this point, 1 ml aliquots were removed and centrifuged. The supernatant was removed and cell pellet was immediately frozen at -20 °C. These samples were designated as the Zero time points. To the remaining culture volumes, varying amounts of arabinose were added to obtain the following final concentrations: 0.2%, 0.02%, 0.002%, 0.0002% and 0.00002%. Incubation was continued at 37 °C with shaking. At set time points, 1 ml aliquots were removed and treated in the same manner as the zero time point.

SDS-PAGE gels were run to analyze all the samples collected (figure 11). Cell pellets were first resuspended in 100 μl of 1x SDS-PAGE sample buffer (Sambrook *et al.*). The samples were heated to 95 °C for 5 minutes and loaded on to the gel. The gel was run in 1x Tris-Glycine buffer at 20 amp for 4 hours. After electrophoresis, the gels were stained with Coomassie blue and the band of increasing intensity at various arabinose concentrations was identified.

To determine the length of arabinose induction, overnight cultures were subcultured 1:100 and incubated until the OD_{600} was 0.6-0.8. A 1:4 subculture was done, two sets were maintained, one with 0.2% arabinose and other without. Aliquots were taken at the following time points 0, 1, 2, 4 and 18 hours. 12.5% SDS-PAGE gels were run and proteins were detected by staining with Coomassie blue (figure12).

Western blotting was performed on some protein gels to confirm the presence of HsdM protein as first described by Towbin *et al* (1979) and as modified by the Biorad electroblot apparatus directions. The samples were run on 12.5% SDS PAGE and were electroblotted on to PVDF membrane (0.2 micrometer). The blotted gel was stained in Coomassie blue (figure 17).

The ECL western blotting from Amersham Pharmacia Biotech was used for detection of immobilized specific antigens is a light emitting non-radioactive method, employing horseradish peroxidase-labeled antibodies. The immunodetection was done as follows. After electrophoresis and blotting, the membrane was blocked by incubation in 5% powdered milk in PBS (pH 7.5, 80 mM Na_2HPO_4, 20 mM $NaH_2PO_4.2H_2O$, 100 mM NaCl) for overnight on an orbital shaker. The blot was then washed extensively in PBS-T (PBS+ 0.1% Tween 20; 2 washes, 15 minutes each). Polyclonal antibody raised against *Eco*R124 endonuclease protein was diluted in 5% milk-PBS and added to membrane. The blot was incubated for 1 hour at room temperature in this primary antibody solution and then washed as before. Horseradish-peroxidase-labeled antirabbit antibody in PBS-T was added to the membrane and incubated for 1 hour at room temperature. The washing was done as before. Then detection was done as follows. Equal volumes of ECL detection solution 1 were mixed with detection solution 2 to give sufficient cover to membrane. The detection

reagent was added to membrane and incubated for 1 minute at room temperature without agitation. The excess detection reagent was drained off and membrane was wrapped in saran wrap making sure it was smooth. The blot was placed in the film cassette; a sheet of autoradiography film was placed on top of membrane and exposed for varying lengths of time (15 seconds- 5 minutes).

Protein Purification

Recombinant methyltransferase protein was isolated from LMG194 cells carrying the pSRI5 plasmid. The purification protocol was adapted from Ramsden and Dreier (1996). Cells were grown overnight in 2 liters LB with 100 μg/ml ampicillin + 0.2% arabinose. The cells were harvested by centrifugation (5000 xg, 4°C, 15 min, Sorvall GSA rotor), resuspended in buffer A (20 mM Tris.HCl, pH 8.0; 50 mM NaCl; 5 mM $MgCl_2$; 7 mM 2-Mercaptoethanol) and lysed by sonication (Branson Model 250 Sonicator). The following settings were used: low power, 45% time on, output control 7 for 2 min and then cool for 3 min. The process was repeated 4-5 times. The extract was kept on ice during the entire process of purification. The disrupted cell sample was centrifuged at 4 °C, 10,000 xg for 15 minutes (SS34 rotor; Sorvall centrifuge). The supernatant was transferred to an ultracentrifuge tube and was centrifuged at 4 °C, 30,000 xg in a Ti70 rotor, Beckman L8-M ultracentrifuge for 2 hours. The supernatant was transferred to a beaker and the volume was noted to precipitate nucleic acid; the NaCl concentration was adjusted to 0.4 M and Polyethelenimine was added to a final concentration of 0.4%. After being stirred at 4 °C for 15 min, the sample was centrifuged at 4 °C, 10,000 xg for 15 minutes. The proteins in the supernatant were precipitated by slowly adding solid ammonium sulfate to a final concentration of 70% (w/v) saturation with stirring. The pellet obtained after centrifugation

(4 °C, 10,000 xg, 30 min) was resuspended in buffer B (20 mM Tris-HCl, pH 8.0; 50 mM NaCl; 0.5 mM EDTA; 7 mM 2-Mercaptoethanol) and dialyzed against the same buffer. After dialysis, the sample (7ml) was applied to a 20 ml DEAE- Sephacel column (Amersham Pharmacia Biotech), washed with buffer B and eluted with 0.05-1 M NaCl gradient. After subsequent dialysis, the peak fractions were loaded on to Heparin agarose column (~ 7ml bed volume; Amersham Pharmacia Biotech) and eluted with a 0.05-1M NaCl gradient. For both columns, protein-containing fractions were detected with a UV monitor and protein concentration determined by Biorad Protein Assay. Fractions were analyzed by SDS PAGE and by Western Blot in some cases. HI0216 methyltransferase containing fractions from the Heparin agarose column were pooled and Centricon YM-30 centrifugal filter units (Millipore) were used to concentrate the sample.

***In vitro* DNA methylation assays**

In vitro DNA methylation assays were performed in a solution of 10 mM Tris-HCl, pH 7.5; 50 mM NaCl; 10 mM EDTA; 5 mM 2mercaptoethanol; 1 μM S-adenosyl (methyl-^{3}H) methionine (85Ci/mMol; Amersham Pharmacia Biotech) and 500 μg/ml DNA. (Meister, 1993). Reaction volumes were 50 μl to which 5 μl of purified methyltransferase was added. The DNA template was phage lambda DNA (NEB or Promega) precut with either *Eco*RI and *Hin*DIII or *Pvu*II was added. Reactions were incubated at 37 °C overnight. The reaction mixes were subsequently extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The aqueous, DNA-containing layer was passed over a gel filtration column (Biorad Micro Bio-Spin® P-30) to remove the unincorporated S-adenosyl (methyl-^{3}H) methionine label. The labeled DNA samples were run on a 5% nondenaturing polyacrylamide gel in 1x TBE as described in Sambrook *et al*

(1989). After electrophoresis the gel was stained in 0.5 μg/ml EtBr, photographed and processed for fluorography by soaking in 1 M-sodium salicylate for 1 hour (Chamberlain, 1979). The gel was then dried down with a Gel Dryer (Gel Dryer Model 583 from BioRad).The dried gel was autoradiographed at -70 °C for a minimum of one week prior to film developement.

To determine the sizes of the labeled DNA fragments, R_f values were calculated for all the bands visible in the picture of EtBr stained gel as well as for the bands in the autoradiograph. The R_f values were calculated by first measuring the distance between the well and the DNA band. This value was then divided by the migration distance of the largest DNA fragment in that lane. Such ratios can then be used when comparing the photograph and autoradiograph to identify the labeled bands.

Results

Cloning of the HI0216 gene using PCR

We were interested in cloning the HI0216 putative *hsdS* gene into a suitable expression plasmid, which would allow for regulated gene expression. This plasmid could then be used in conjunction with an *hsd*M and *hsd*R expression system to test the functionality of the HI0216 gene. The vector used was pBAD22 as it has several desirable features (Guzman *et al.*, 1995). This vector contains the tightly regulated promoter, P_{BAD}, whose expression is induced by the addition of arabinose to the medium. An *Nco*I site (CCATGG) is located at the correct distance downstream of the ribosome binding site (RBS) such that the ATG in the *Nco*I site can serve as the translation start codon for polypeptide biosynthesis (figure 4A). A unique *Xba*I site is located downstream in the multiple cloning site (MCS). The source of the HI0216 gene was pUC18:GHIJT88. Constructed during the *Haemophilus* genomic sequencing project (Fleischmann *et al.*, 1995), pUC18: GHIJT88 contains a 1909 bp *H. influenza* fragment which includes the entire HI0216 open reading frame (The Institute for Genomic Research; American Type Culture Collection).

In order to introduce a *Nco*I site, to overlap the HI0216 start codon, the 5' end of the gene was PCR amplified. Two oligonucleotides served as primers for the reaction carried out by the thermostable Vent DNA polymerase (New England Biolabs). The 5' oligonucleotide overlaps the start codon and includes the new *Nco*I site (figure 3). The 3' oligonucleotide is complementary to an internal region of *hsdS* that includes an *Xba*I site. The PCR reaction produced a fragment of predicted size (230 bp) as determined by agarose gel electrophoresis (figure 5)

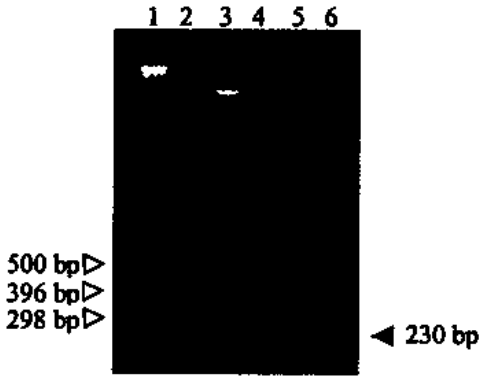


Figure 5. Agarose gel electrophoresis of PCR product.
The 5' end of the *hsd*S gene was PCR amplified to introduce a 5' *Nco*I site the resulting PCR product (lane 6) was estimated to be 230 bp. Lane 1, Molecular weight marker of indicated sizes (open arrows); Lanes 2, 3, 4, N/A; Lane 5, Negative Control; Lane 6, PCR product 230 bp.

The PCR product was digested with *Nco*I and *Xba*I and mixed with pBAD22 digested with the same enzymes. After ligation of the DNAs and transformation into a suitable host, the resulting plasmid pSRI1 was identified by digestion with *Ssp*I (figure 6). DNA bands obtained at 0.8 Kb and 0.6 Kb showed that the PCR insert was present.

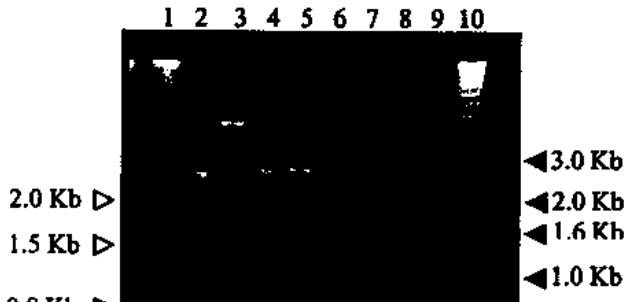


Figure 6. pSRI1 isolates digested with *Ssp*I to confirm PCR insert.
Lanes 1 and 10, Molecular weight marker of indicated sizes (block arrows); Lane 2, Uncut pBAD22; Lane 3, *Ssp*I digested pBAD22; Lanes 4,6,8, Uncut isolates of pSRI1; Lanes 5,7,9, Digested miniprep DNAs; Lanes 7,9, Correct pSRI4 constructs with *Ssp*I fragments of 2.0 Kb, 1.5 Kb, 0.8 Kb, 0.6 Kb (open arrows)

Subcloning of the rest of *hsd*S gene was performed as described in Materials and Methods (also see figure 4B). The resulting cloned plasmids had two possible orientations due to the fact that there were *Xba*I sites at both ends of the insert. To confirm the correct orientation plasmid, *Eco*RI was used to digest DNA. The correctly oriented plasmid produced two fragments with sizes 0.6 Kb and 5.2 Kb. The incorrect orientation produced a 0.9 Kb and a 4.9 Kb *Eco*RI fragments (figure 7). The resulting plasmid, pSRI4, is 5.8 Kb long and has the Ampicillin resistance gene, the *ara*C regulatory protein gene and the entire HI0216 *hsd*S gene under the control of the P_{BAD} promoter (figure 7A).

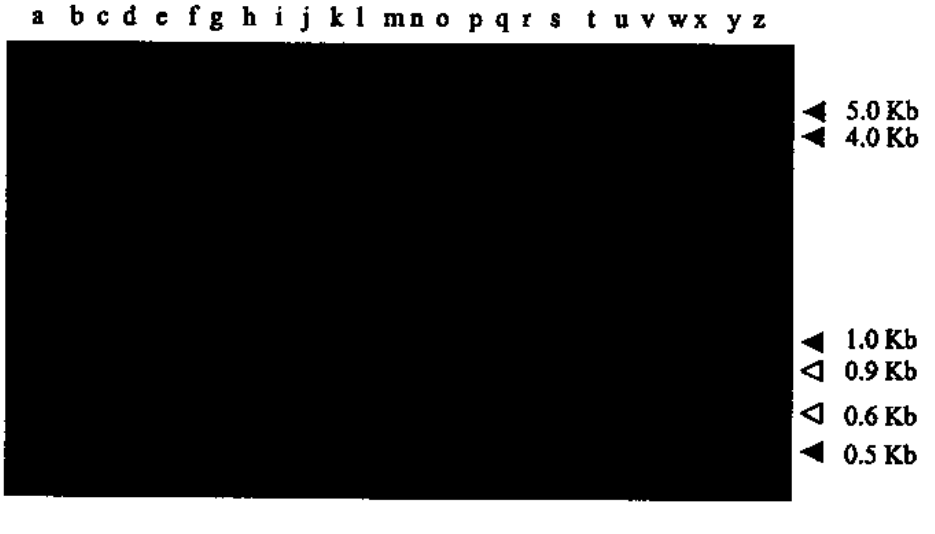


Figure 7. *Eco*RI digestion of pSRI4 isolates to confirm orientation.
Lane a, Molecular weight marker of indicated sizes (block arrows); Lanes b, c, Uncut pBAD22; Lanes d-m, Uncut miniprep DNAs; Lanes n,o, Digested pBAD22; Lanes p-y, Digested DNA isolates; Lanes q,s,v,y, Isolates showing incorrect orientation with 0.9 Kb and 4.0 Kb; Lanes p,r,u, Isolates showing correct orientation with 0.6 Kb and 5.2 Kb. The correct orientation plasmid was called pSRI4.

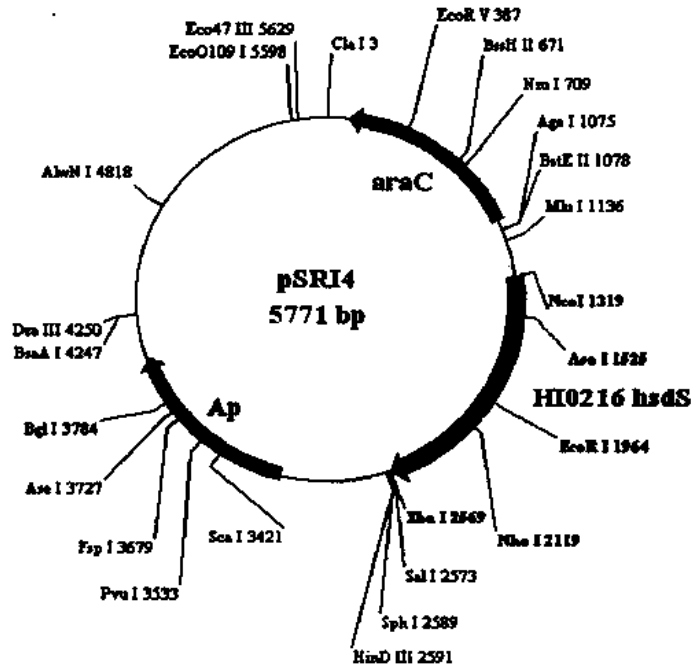


Figure. 7A. pSRI4 plasmid map showing the *hsd*S gene insert from *H. influenzae*.

Construction of a HI0216 Methyltransferase expression vector

The plasmid, pSRI5 (figure 8), was constructed to be used as an expression vector for HI0216 methyltransferase purification. pSRI5 is a derivative of pSRI4 in which the *Eco*DXXI *hsd*M gene from pMMW56 was inserted into the *Sal*I-*Bam*HI sites of pSRI4 (see Materials and Methods). The *hsd*M insert introduces a second *Bsp*DI (isoschizomer of *Cla*I) into the plasmid. Thus prospective positive transformants were screened by *Bsp*DI digestion of miniprep DNA. (figure 9)

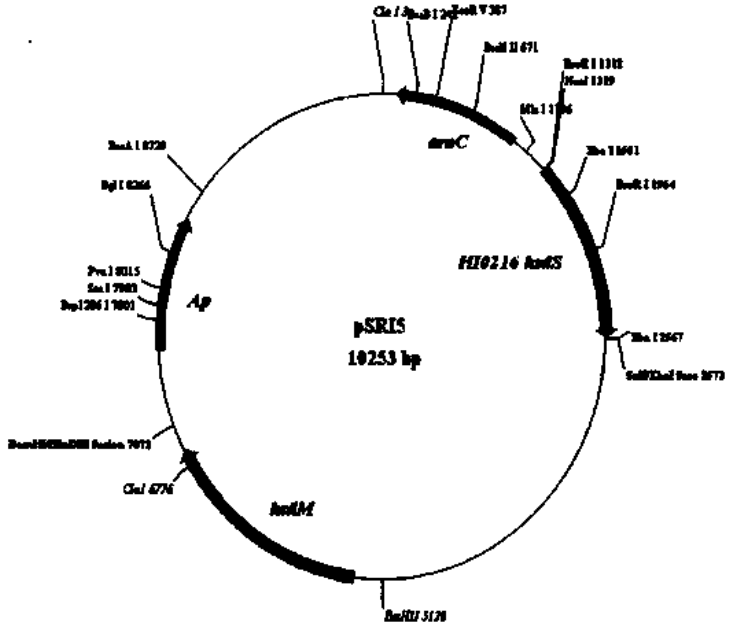


Figure 8. Plasmid map of pSRI5 showing *hsd*S and *hsd*M inserts.

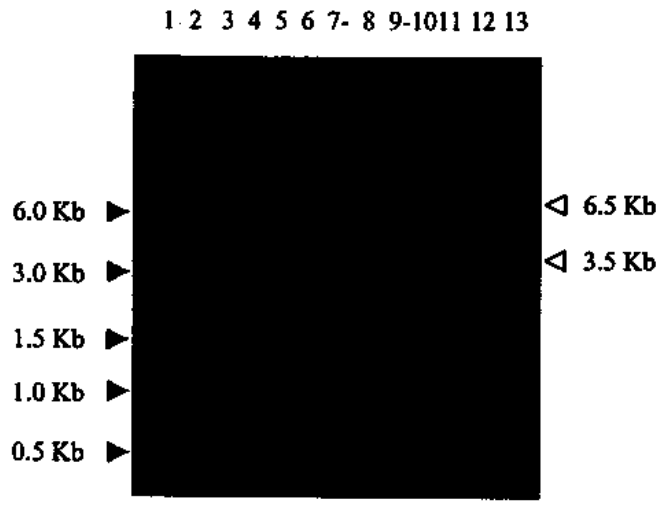


Figure 9.Confirmation of pSRI5 inserts by digestion with *Bsp*DI (*Cla*I).
Lane 1, Molecular weight marker of indicated sizes (block arrows); Lane 2, Uncut pSRI4; Lane 3, Uncut pMMW56; Lanes 4-7, Uncut isolates; Lane 8, Digested pSRI4; Lane 9, Digested pMMW56; Lanes 10-13, Digested pSRI 5 isolates showing the 6.5 Kb and 3.5 Kb fragments (open arrows).

In vivo restriction assay

The *hsd*S construct, pSRI4, was analyzed for its ability to provide a functional type IC specificity subunit when coexpressed in LMG194 cells with the *hsd*M and *hsd*R genes from pHJ8. The HI0216 *hsd*S gene in pSRI4 is under the control of the P_{BAD} promoter and thus its expression could be regulated by the addition of arabinose to the medium. The *in vivo* restriction assay used monitored a bacterial strain's ability to inhibit (restrict) the growth of bacteriophage λ as compared to a nonrestricting control strain. The efficiency of plating (EOP) was used as a measure of restriction activity (figure 10). An EOP value of much less than 1 indicates restriction activity while an EOP of approximately 1 demonstrates the lack of restriction. The results are shown in table VI.

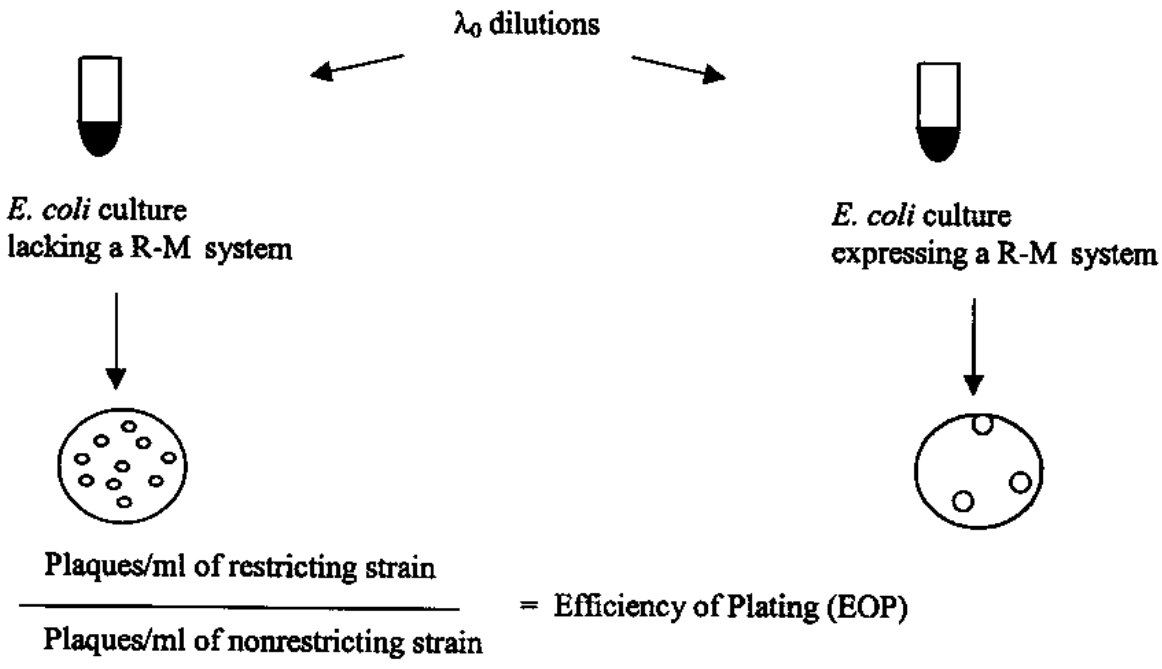


Figure 10. Protocol for *in vivo* restriction assay.

LMG194 /pHJ8 cells were used as the control, nonrestricting strain and thus, by definition had an EOP of 1 (Table VI). The presence of pSRI4 alone did not restrict phage λ regardless of whether or not *hsd*S expression was induced by the addition of arabinose to the medium. In the absence of arabinose, LMG194 cells containing both pHJ8 and pSRI4 had an EOP of 1.03 and thus did not restrict phage growth. On the other hand, induction of HI0216 *hsd*S expression by the addition of arabinose to LMG194/pHJ8+pSRI4 led to a drastic reduction in the EOP value. Thus the induced HI0216 HsdS polypeptide could assemble with the constitutively produced HsdM and HsdR subunits from pHJ8 to produce a functional restriction endonuclease. The ability of the HI0216 gene to complement the enteric system demonstrated that HI0216 is a member of type IC family of R-M systems.

Table VI. EOP values of restriction assays using LMG194 cells containing pSRI4 and/or pHJ8.

Plasmids	*hsd* genes	Arabinose Added	EOP
pSRI4	*hsd*S (HI0216)	+ –	1.6 0.8
pHJ8	*hsd*R, *hsd*M	+ –	1.0 1.0
pSRI4+pHJ8	*hsd*S+*hsd*R, *hsd*M	+ –	5.4×10^{-3} 1.03

To verify that the plasmid pSRI5 (constructed to overexpress the HI0216 methyltransferase for protein purification) produced a functional product, it was tested in the *in vivo* restriction assay along with the HsdR producing plasmid, pMG3. The results are shown in table VII. LMG194/pSRI5+pMG3 cells preincubated with arabinose produced a functional restriction endonuclease (EOP=0.089) while the same strain did not restrict λ in the absence of arabinose. Bacteria containing either of the plasmids alone were restriction minus. These results demonstrated the HsdM and HsdS subunits expressed from pSRI5 could assemble into a functional enzyme.

Table VII. Restriction assay EOP values using LMG194 cells containing pSRI5 and/or pMG3.

Plasmid	*hsd* genes	Arabinose added	EOP
pSRI5	*hsd*S, *hsd*M	+	0.90
		–	0.63
pMG3	*hsd*R	+	1.0
		–	1.0
pSRI5+pMG3	*hsd*S, *hsd*M+ *hsd*R	+	8.89×10^{-2}
		–	0.8

Protein expression optimization

In protein expression optimization studies, mid-log cultures of LMG194/pSRI5 were induced for four hours with varying concentrations of arabinose to check at which concentration the HsdS protein expression was optimal. HsdS production was monitored by SDS-PAGE (Sambrook *et al.*, 1989) followed by Coomassie blue staining of the 12.5% polyacrylamide gel (figure 11). Arabinose concentrations below 0.0002% did not result in a visible HsdS band (lane 6), while an induced band of similar intensity was present at concentrations of 0.2, 0.02 and 0.002% arabinose (lanes 1-3).

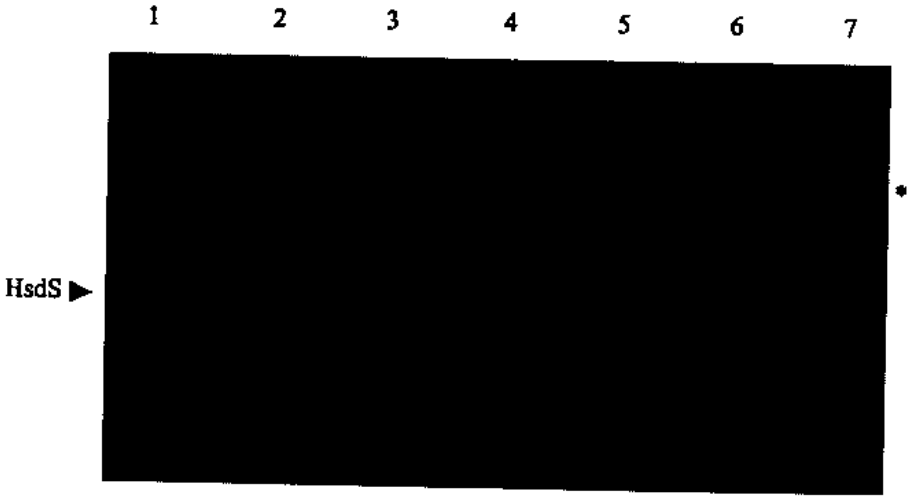


Figure. 11. Protein expression induction at various arabinose concentrations.
Mid-log cultures of LMG194/pSRI5 were induced with varying concentrations of arabinose for 4 hours. 1ml aliquots were centrifuged and the cell pellets were resuspended in gel sample buffer. After being heated to 95°C, the denatured cell lysates were analyzed by SDS-PAGE on a 12.5% polyacrylamide gel.
Lane 7, Induction positive control (* indicates Myc-His-LacZ fusion protein); Lanes 1-4,6, are LMG194/pSRI5 lysates induced with the following arabinose concentrations: Lane 1, 0.2%; Lane 2, 0.02%; Lane 3, 0.002%; Lane 4, 0.0002%; Lane 6, 0.00002%; Arrow indicates the ~45kDa HsdS band.

To determine the optimal induction time period, LMG194/pSRI5 cells were incubated with 0.2% arabinose for varying lengths of time. 1 ml aliquots were removed at indicated time points (figure 12) and processed as described for the inducer optimization experiments. The results are shown in figure 12. From the 4 hour time point onwards, high levels of HsdS protein expression were achieved. The largest total amount of HsdS was seen in the overnight sample (figure 12, lane 5).

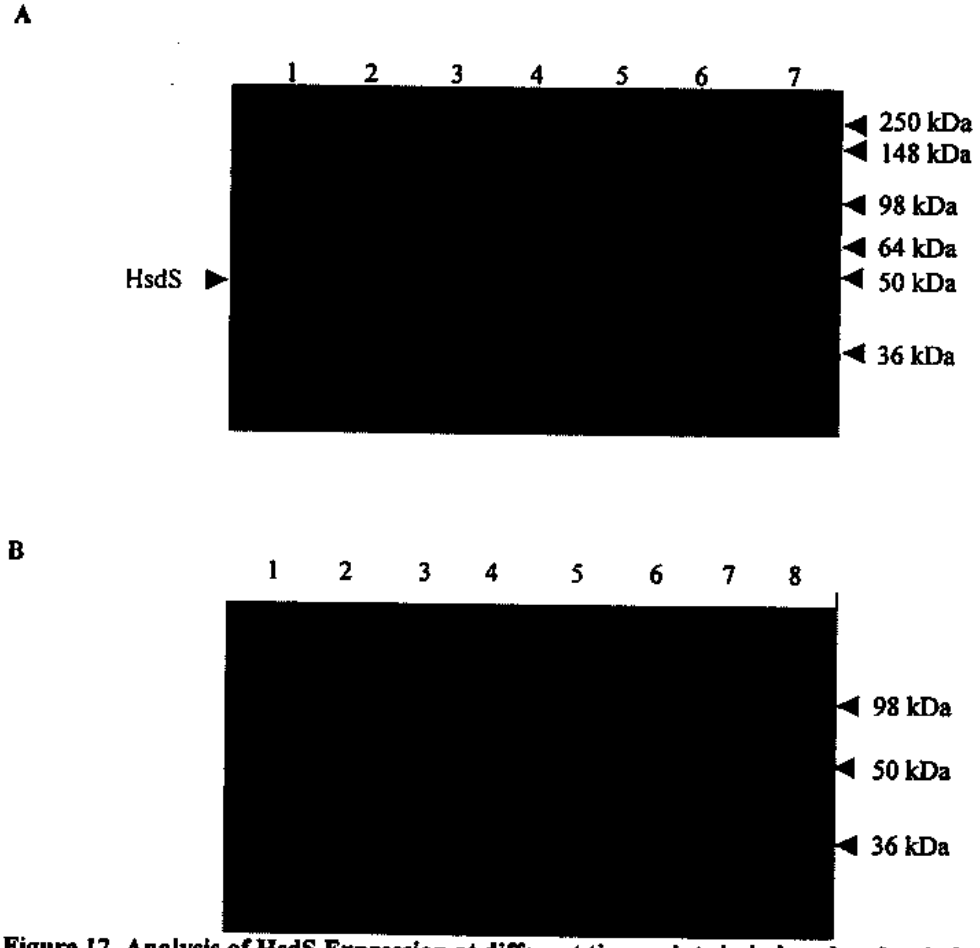


Figure 12. Analysis of HsdS Expression at different time points in induced and uninduced cultures of LMG194/pSRI5.

LMG194/pSRI5 cultures were grown under inducing (A; 0.2% arabinose) or noninducing (B) conditions. At various times after induction, 1ml aliquots were removed and processed as described for Figure 11. Samples were then analyzed on 12% polyacrylamide gels under denaturing conditions.

Gel A. Induced culture sampled at the following time points: Lane 1, 0 hour; Lane 2, 1 hour; Lane 3, 2 hours; Lane 4, 4 hours; Lane 5, 18 hours (overnight); Lane 7, Molecular weight marker; Lane 6, not relevant.

Gel B. Uninduced culture sampled at the following time points: Lane 1, 0 hour; Lane 2, 1 hour; Lane 3, 2 hours; Lane 5, 4 hours; Lane 6, 18 hours (overnight); Lane 7, Positive control; Lane 8, Molecular weight marker of following sizes in kDa: 250,148,98,64,50, 36 and 22; Lane 4, not relevant. Note the absence of the 42.3kDa sized band in the uninduced cultures.

To monitor the level of HsdS production relative to the total protein present in the cell, cultures aliquots were taken at various times and the samples were processed such that equivalent amounts of protein were loaded in each gel lane (figure 13). From this analysis, it appears that the relative amount of HsdS to total protein remains constant for the 4 hour and overnight samples (figure 13, lanes 3 and 4).

A consideration of the above optimization studies led us to decide on cell growth conditions for protein purification. An overnight induction period was chosen because relative HsdS protein expression was optimal and the total amount of HsdS was the greatest at this time point. An inducer concentration of 0.2% was chosen because this amount well beyond the minimum necessary for maximal HsdS induction.

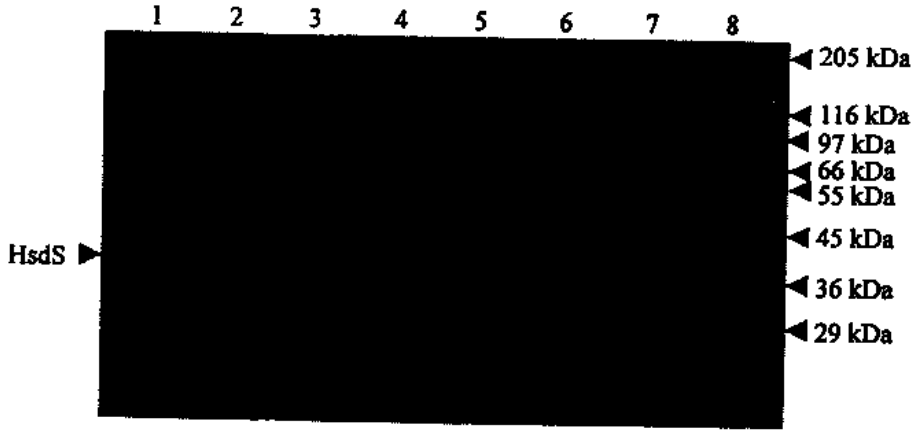


Figure 13. Optimization of Protein Expression using same cell mass of LMG194/pSRI5.
Lanes 1-4, Induced pSRI5 at different time points, Lane 1, 1hour; Lane 2, 2 hours; Lane 3, 4 hours; Lane 4, Overnight; Lanes 5-7, Uninduced pSRI5 at various time points; Lane 5, Overnight; Lane 6, 4 hours; Lane 7, 2 hours; Lane 8, Molecular weight marker of following sizes in kDa: 205, 116, 97, 66,55, 45, 36, 29.

Protein Purification

As a first step towards the further characterization of the hybrid HI0216 DNA methyltransferase, the protein was purified using our pSRI5 overexpression system. Two liters of LMG194/pSRI5 cells were grown to mid-log phase and induced for *hsd*S expression using the conditions determined in the optimization experiments (see previous section). Protein purification was performed as described in Materials and Methods. Purification progress was monitored by SDS-PAGE. Two chromatography columns (DEAE sepacryl and Heparin agarose) were critical for the purification of the methyltransferase. DEAE is a weak anion exchanger and it binds to negatively charged regions of proteins. A partially purified protein sample was loaded on to the DEAE column at low sodium chloride concentration (20 mM) and after extensive column washing, the bound proteins were eluted with a 20 mM-1 M sodium chloride gradient. Figure 14 shows a denaturating 7.5% polyacrylamide gel of the peak protein fractions.

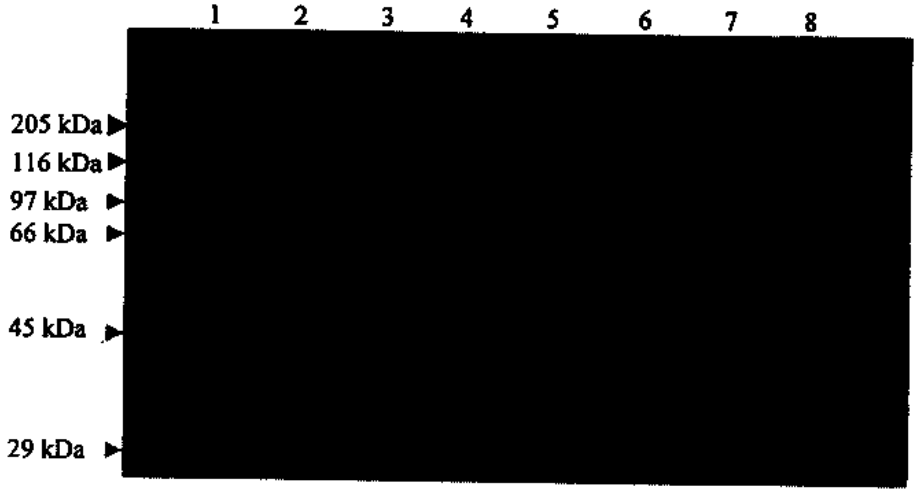


Figure. 14. SDS PAGE analysis of DEAE column fractions from peak of protein elution.
Lane 1, Molecular weight marker of sizes in kDa 205, 116, 97,66,45,29; Lanes 2-8, fractions 14-20 respectively

DEAE fractions 18-21 were pooled dialyzed and loaded onto the Heparin agarose column under low salt conditions (20 mM NaCl). A highly negatively charged molecule, heparin is often used for purification of basic proteins such as DNA binding proteins. A NaCl gradient (20 mM-1 M) was used to elute the bound proteins. The peak protein fractions were run on denaturing 7.5% polyacrylamide gel using the Pharmacia Phast system (figure 15). The HsdM and HsdS bands are indicated in the figure 15. A more extensive SDS-PAGE analysis of column fractions was performed using the Biorad minigel system (figure 16).

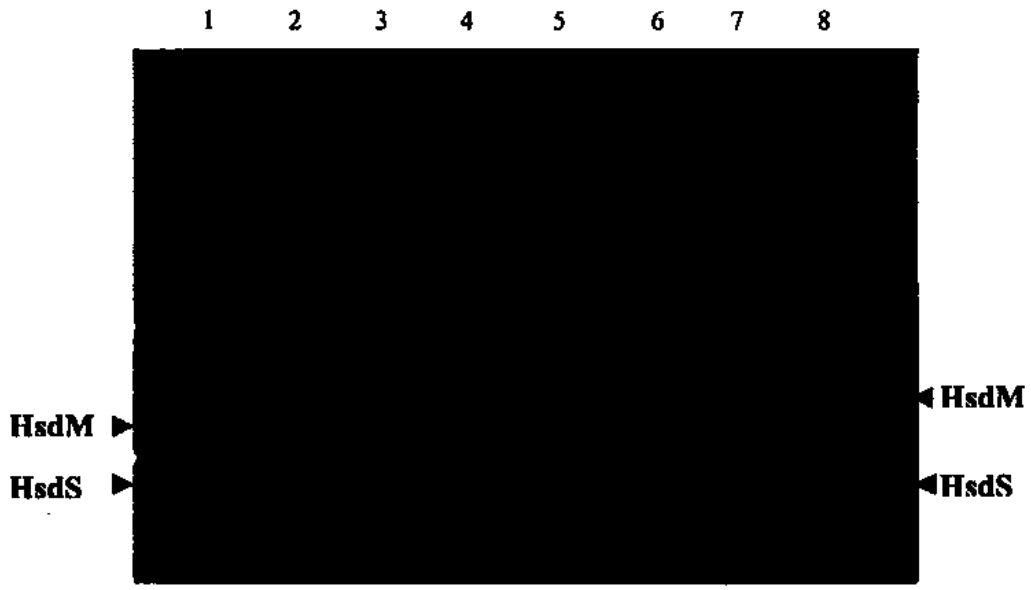
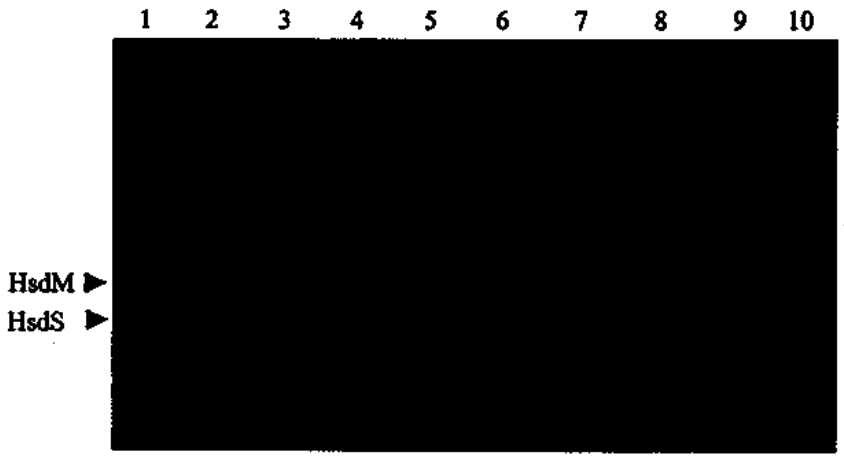


Figure 15. SDS-PAGE analysis of Heparin agarose column fractions from peak of protein elution. Lane 1, Heparin fraction 16; Lane 2, fraction 18; Lane 3, fraction 19; Lane 4, fraction 20; Lane 5, fraction 21; Lane 6, fraction 22; Lane 7, Molecular weight marker; Lane 8, Positive control *Eco*R124II methyltransferase.

A



B

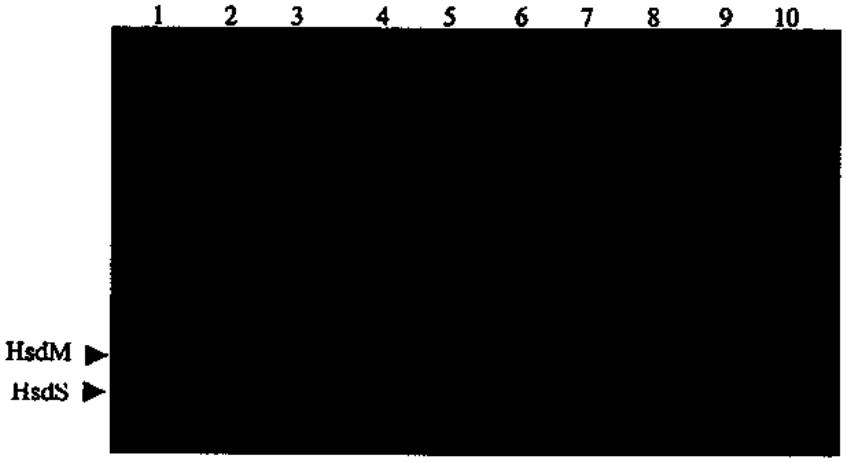


Figure.16. Heparin agarose protein fractions 14-29 on 12.5% SDS PAGE.
Gel (A). Lane 1, Heparin load; Lanes 2-9, HA fractions 14-21, respectively; Lane 10, Molecular weight marker of sizes in kDa 205, 119, 97,66, 45, 29. The HsdM band is found at ~ 50kDa and the HsdS was found at ~ 42.3kDa.
Gel (B). Lanes 1-3, fractions 22-24 respectively; Lane 4, Molecular weight marker; Lane 5-9, fractions25-29 respectively; Lane 10, Positive control purified *Eco*R124 methyltransferase. The HsdM band is found at ~50kDa and the HsdS was found at ~42.3kDa.

The fractions analyzed from heparin agarose columns were pooled together. The fractions pooled were called Pools A, B and C. Pool A consisted of heparin fraction 15 to 18; Pool B consisted of fractions 19 to 22; while the Pool C consisted of fractions 24 to 28. The dividing points between pools were chosen based on the protein band patterns on SDS-PAGE gels. Fractions with similar patterns and band with similar intensities were matched together. These pools were concentrated using centricon filter units. The fractions were stabilized by addition of glycerol to 50% and storage at - 20 °C. These pools were later used in *in vitro* methylation assay.

Western blotting was performed to verify the presence of the HI0216 methyltransferase. The primary antibody used was raised against the *Eco*R124I restriction endonuclease and recognizes the HsdM and HsdR subunits only. The results can be seen in figure 17.

In Lane 2, the DEAE sample produces three bands seen in both the 5 seconds and 5 minutes exposures Lane 3, the heparin load, also contains three bands. Lane 4 contains the Heparin flowthrough which is not bound to the column, Only the middle antibody-tagged protein is present here. Lanes 5, 6 each contain a faint band and lane 7 contains an intense band all of which are of the approximate size of the HsdM positive control in lane 10. Lane 9 shows that the heparin column protein fraction number 22 also has the *hsd*M band.

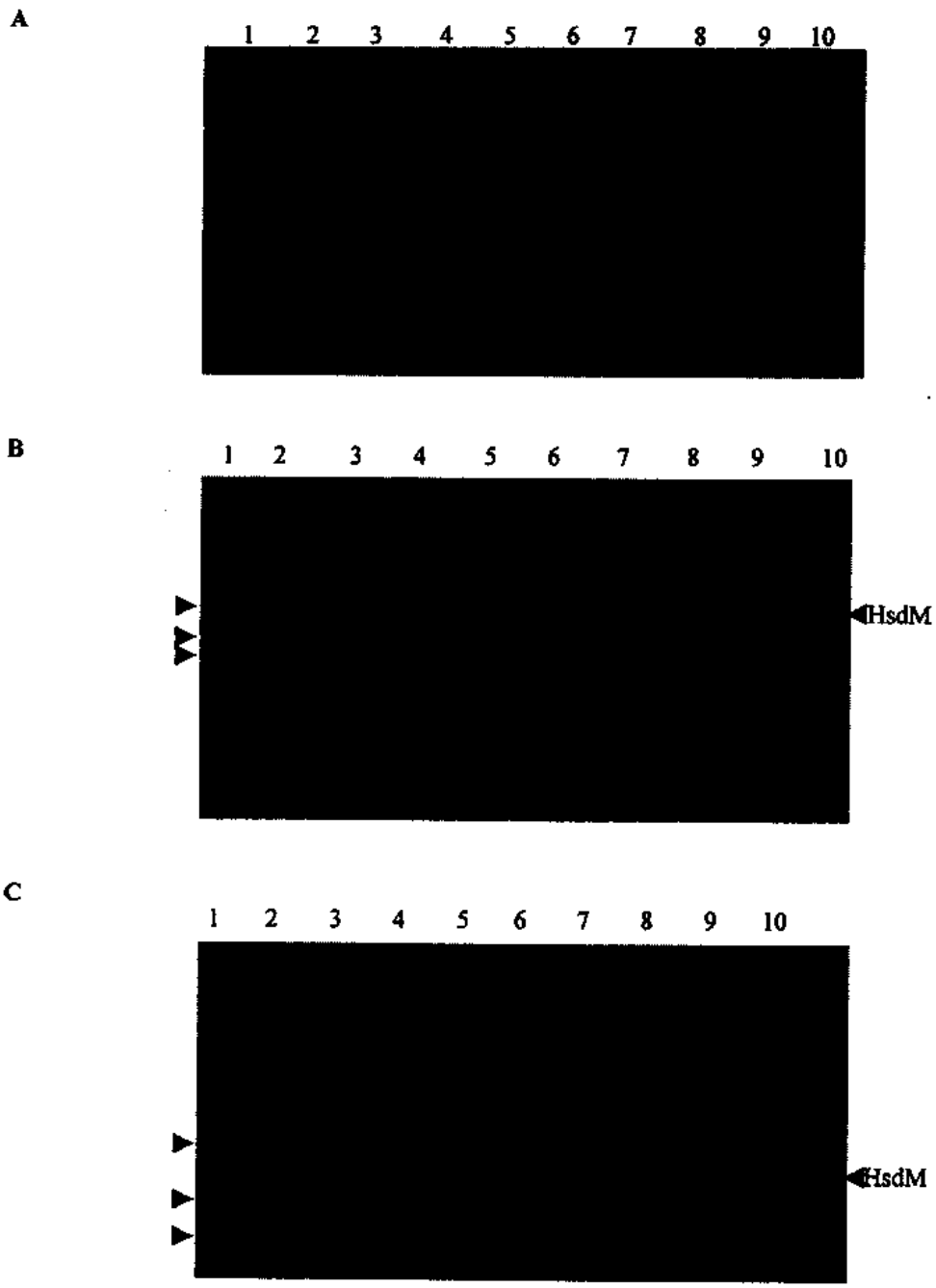


Figure. 17. Confirmation of *hsd*M by Western Blotting Technique using ECL detection kit.
(A) 12.5%SDS Polyacrylamide gel that was electroblotted to membrane
17 B and C shows film of the blot exposed at (B) 5 seconds and (C) 5 minutes
Figures A, B, C: Lane 1, Molecular weight marker of sizes in kDa 205, 119, 97,66,45 and 29.
Lane 2, DEAE load, Lane 3, Heparin Load, Lane 4, Heparin flow through, Lane 5, HA Pool A, Lane 6, HA Pool B, Lane 7,HA Pool C, Lane 8, HA fraction 16, Lane 9, HA.fraction 22, Lane 10, Purified *Eco*R124I methyltransferase.

***In vitro* methylation assay**

In vitro methylation assays were performed in an effort to identify the HI0216 recognition sequence. In these assays, S-adenosyl (methyl ^{3}H) methionine was used as methyl donor so that the DNA sites became radioactively labeled upon modification by the purified methyltransferase. Bacteriophage λ DNA digested with either *Pvu*II or *Hind*III+*Eco*RI was used as the substrate. The fact that λ phage infection was restricted *in vivo* by the HI0216 hybrid restriction system indicates that λ DNA contains one or more HI0216 recognition sites and thus is a good template for the *in vitro* methylation reaction. The DNA reaction products were separated on a 5% nondenaturing polyacrylamide gel. Ethidium bromide staining of the gel was used to visualize all the DNA fragments from the digest (figure 18A), while autoradiography reveals which of the fragments had been radioactively labeled by the enzyme (figure 18B).

The complete sequence of λ is known. From this sequence information, the cleavage sites of the type II restriction enzymes were determined and the sizes of the resulting fragments were calculated. Only a partial digestion was achieved with *Pvu*II and thus the penultimate-sized band (8.4 Kb) is a partial digest fragment. The smaller sized *Pvu*II fragments are not visible in this photo.

*Eco*RI DNA methyltransferase (New England Biolabs) was used as a positive control for DNA methylation. Apparently this enzyme was used in excess because all the λ*Pvu*II DNA fragments were labeled (figure 18B, lanes 3, 4). No methylation of λ DNA incubated with Heparin agarose (HA) poolB was detectable (figure 18B, lanes 5 and 6). Incubation of λ DNA with HA poolA resulted in the methylation of at least three λ *Eco*RI-*Hind*III fragments, 21.2 Kb, 4.9 Kb and 1.9 Kb (figure 18B, lane 8). Atleast two λ *Pvu*II fragments

(8.4 Kb and 3.9 Kb) were also labeled by HA pool A (lane 7). To identify the actual recognition site, the DNA sequences of these fragments will need to be compared to find an identical sequence which also matches pattern of type IC enzyme recognition ($XXAN_{6-8}RTXX$ where X= a specific nucleotide, R= G or A, and N= any base). Further experimentation is required to definitively identify the recognition site.

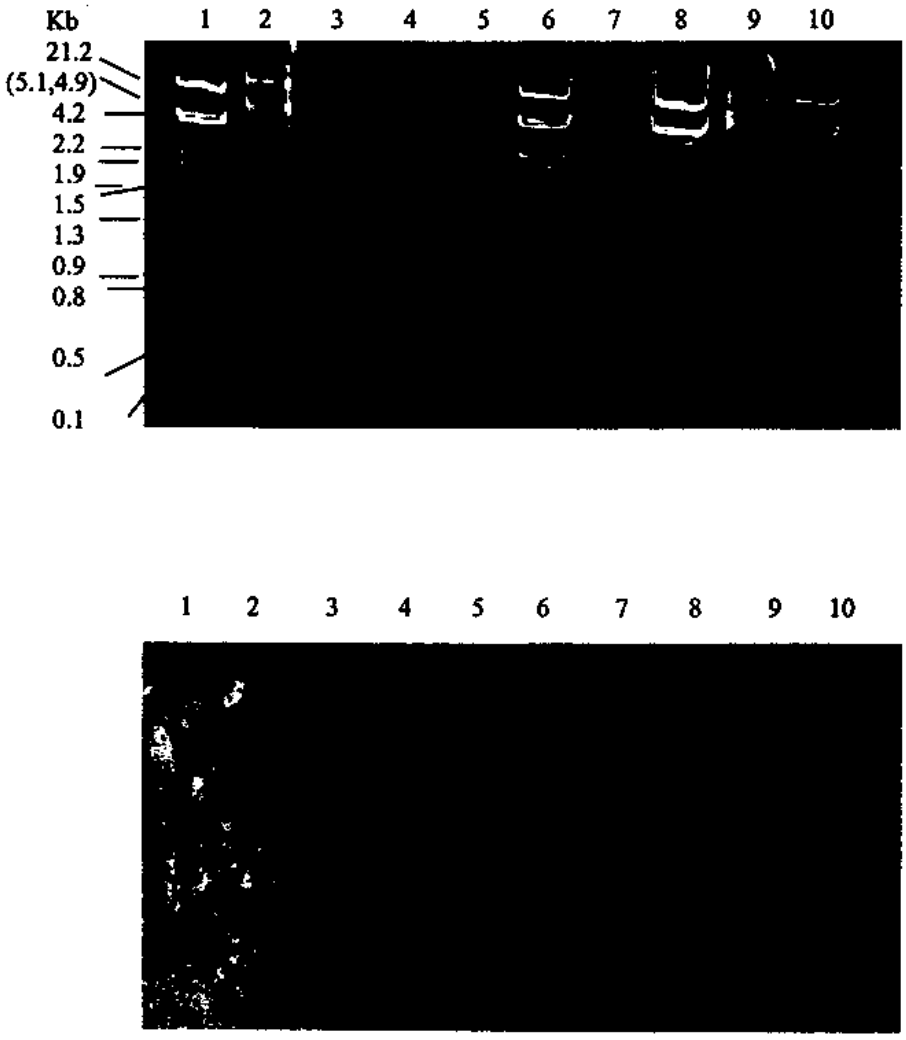


Figure. 18. *In vitro* DNA methylation assay.
(A) 5% acrylamide gel electrophoresis of reaction mixture, (B) autoradiograph of the gel in A.
Gel A and B : Lanes 1and 10; λ*Hind*III-*Eco*RI molecular weight markers (Promega); Lanes 2 and 9, Negative control λ*Eco*RI+*Hind*III with no enzyme added; Lane 3 and 4, Positive control, λ*Pvu*II-*Eco*RI methyltransferase; Lane 5,. λ*Pvu*II- HA pool B; Lane 6, λ*Eco*RI+*Hind*III- HA Pool B; Lane 7, λPvuII-HA Pool A; Lane 8, λ*Eco*RI-*Hind*III-HA Pool A.

Discussion

R-M systems serve as barriers to protect the host from foreign DNA and viruses. The ability to distinguish between "self" DNA and "nonself" DNA is due to methylation of the host at specific residues within the recognition sequence for the R-M system(Redaschi and Bickle, 1996). Type I systems recognize asymmetric sequences and if sequence is unmethylated they cut the DNA at random positions (Bickle and Krüger, 1993; Murray, 2000). The type I systems are suited to a flexible defense strategy because the HsdS subunit is a separate entity from both methylation and restriction units. When assembled into a complex, HsdS serves as DNA specificity determinant for both methylation and restriction activities of the complex. Thus a single point mutation can theoretically generate a new recognition sequence specificity for the R-M system. This is advantageous because at a low frequency, the foreign viral DNA may escape restriction and become modified. The subsequent viral progeny would then be immune to the restriction endonuclease. Evolution of a small, subpopulation of cells with a novel recognition pattern will ensure the survival of these organisms.

In contrast, type II systems are composed of separate enzymes, which carry out their methylation and restriction activities independently. Each specifies an identical recognition sequence. At least two different mutations would be required for the emergence of a new type II system specificity. Such a double mutation would be quite a rare event: if the rate of mutation for a single mutation is approximately 1 in 10^8 cells, then a double mutant would only arise approximately 1 in 10^{16} (Maloy *et al.*, 1994; Martinez and Baquero, 2000). The type IC HsdS subunits appear to have been designed for ease of generation of new binding

specificities. The HsdS subunits are interchangeable within a family resulting in hybrid but functional proteins (Redaschi and Bickle, 1996). Sequence comparisons amongst HsdS subunits reveal the type IC family members share three regions of sequence similarity separated by two regions that vary considerably between individuals. The variable domains are responsible for DNA recognition while the constant regions are thought to be responsible for protein-protein interactions with the HsdM and HsdR subunits. DNA recombination between the central conserved regions can yield a hybrid *hsd*S gene which recognizes a sequence that is a hybrid of the two parental systems (Gubler *et al.*, 1992). In addition, the central, conserved region has the tetraaminoacid repeat which governs the length of the nonspecific spacer in the DNA recognition sequence. Variation in the number of repeats leads to a change in the spacer distance in the recognition sequence (Price *et al.*, 1989); (Gubler and Bickle, 1991). In this study we wanted to clone and characterize the HI0216 *hsd*S gene from *H. influenzae.* Based on the amino acid sequence comparisons, HI0216 appeared to be a member of the type IC family (figure 17).

Figure.19. Amino acid sequence comparison between HI0216 and *Eco*DXXI *hsd*S gene.
Sequence comparison was performed using the Blast-2 sequences program [(Tatusova and Madden, 1999); http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html]

The three conserved regions of the enteric (Type IC) *Eco*DXXI *hsd*S subunit were quite similar to the equivalent regions of HI0216; however there was little similarity in the variable regions. In addition, HI0216 has an additional tetraamino acid repeat in the central conserved region that might yield a novel spacer length in the recognition sequence.

To demonstrate that HI0216 encoded a functional HsdS, we constructed a hybrid system which consisted of the enteric-derived *hsd*M and *hsd*R genes present on pHJ8 and the inducible *hsd*S gene on pSRI4. *In vivo* restriction assays demonstrated that the HI0216 complemented the enteric system and the heterologous HsdR, HsdM and HsdS subunits could assemble into a functional restriction endonuclease. It was also observed that for restriction to occur, all three subunits were essential. Arabinose preincubation which induced HI0216 expression was required for detectable restriction. Thus HI0216 served as the recognition subunit for this hybrid Type IC system.

In the isolation of methyltransferase protein optimal conditions of arabinose concentrations and overnight incubation resulted in maximal isolation and this was used in *in vitro* DNA methylation reactions to identify the recognition sequence. The methyltransferase protein consists of HsdM and HsdS subunits and the bands were identified in SDS PAGE gels with the HsdM running at approximately at 55 kDa and HsdS at approximately 42.3 kDa.

In the identification of recognition sites using *in vitro* methylation reactions, the DNA fragments containing the HI0216 recognition site were were methylated and radioactively labeled by the enzyme preparation. A future direction will be to locate a common sequence present in all the labeled DNA fragments, which conforms to the type IC site organization. *In vitro* methylation assays employing double-stranded oligonucleotides containing derivatives of the sequence will help confirm the identity of the HI0216 binding site.

Our results show that the type IC R-M systems are not restricted to the enteric genuses of *Escherichia* and *Salmonella* but are also found in their distant cousin, *Haemophilus*. The ability of subunits from different species to interact suggests a strong conservation of amino acid sequence and function during evolution. This conservation is restricted to the subunit interaction domains while the DNA binding domains appear to be free to diversify. Another noteable feature is that the HI0216 gene and its partner *hsdM* and *hsdR* genes have a chromosomal location. Thus the previous classification of type IC enzymes as having an episomally-based origin was merely due to the small number of members within the family.

Literature Cited

Arber, W. (1990) Mechanisms in microbial evolution. *J. Struct. Biol.*, **104**, 107-111.

Arber, W. (1994) On the generation of genetic diversity in microorganisms. *Proc. Indian natn. Sci.Acad.*, **B60**, 357-364.

Bickle, T.A. and Krüger, D.H. (1993) The biology of DNA restriction. *Microbiol. Rev.*, **57**, 434-450.

Chamberlain, J.P. (1979). Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate. *Anal Biochem.*, **15**; 98(1):132-5.

Dreier, J., MacWilliams, M.P. and Bickle, T.A. (1996) DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.*, **264**, 722-733.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science.*, **269**, 496-512.

Gubler, M. and Bickle, T.A. (1991) Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.*, **10**, 951-957.

Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A. and Bickle, T.A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.*, **11**, 233-240.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121-4130.

Hanahan, D. (1983). Studies on transformation of *Escherichia.coli* with plasmids. *J. Mol Biol.*, **166**, 557-580.

Maloy, S., Cronan, J. and Freifelder, D. (1994) *Microbial Genetics*. Jones and Bartlett Publishers, Boston.

Martinez, J. and Baquero, F. (2000) Mutation Frequencies and Antibiotic Resistance. *Antimicrob. Agents Chemother.*, **44**, 1771-1777.

Murray, N. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev.*, **64**, 412-434.

Price, C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989) Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 Type I DNA restriction and modification enzymes. *J. Mol. Biol.*, **205**, 115-125.

Ramsden, J. J. Dreier, J. (1996) Kinetics of the interaction between DNA and the type IC restriction enzyme *Eco*R124II. *Biochemistry.*, **26**; 35(12):3746-53.

Redaschi, N. and Bickle, T.A. (1996) DNA restriction and modification systems. In Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Brooks Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, R., Schaechter, M. and Umbarger, H.E. (eds.), *Escherichia coli and Salmonella : Cellular and Molecular Biology, second edition.* ASM Press, Washington, D.C., pp. 773-781.

Silhavy, T. J., Berman, M. L. and Enquist, L. W. (1984). Experiments with Gene Fusions. Cold spring Harbor.

Snyder, L. (1995) Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.*, **15**, 415-420.

Tatusova, T.A. and Madden, T.L. (1999) Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett.*, **172**, 247-250.

Towbin, H. Staehelin, T. Gordon. J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci.*, U S A **76** (9): 4350-4.